ST. LAWRENCE HIGH SCHOOL

A JESUIT CHRISTIAN MINORITY INSTITUTION

FOR GOD AND COUNTRY

STUDY MATERIAL-5
SUBJECT – STATISTICS

Pre-test

Chapter: BIVARIATE ANALYSIS **Class: XII**

Topic:Regression and Rank correlation **Date: 14.05.2020**

REGRESSION

&

RANK CORRELATION

PART 2

Problems on regression and rank correlation:

Q1. Given the regression lines $2x + 3y = 7 \text{ and } 5x + y = 11$.

a. Find mean of x and y.
b. Identify the regression lines.
c. Determine correlation coefficient.
d. Determine the ratio of variances of x and y.

Ans: a. Since the regression lines intersect at $(\bar{x}, \bar{y})$.

Solving the given equations we get $x = 2 \text{ and } y = 1$.

So $\bar{x} = 2 \text{ and } \bar{y} = 1$

b. Let the line $2x + 3y = 7$ is regression equation y on x .

$$2x + 3y = 7 \Rightarrow y = \frac{-2}{3}x + \frac{7}{3} \Rightarrow b_{yx} = \frac{-2}{3}$$

And the line $5x + y = 11$ is regression equation x on y .

$$5x + y = 11 \Rightarrow x = \frac{-1}{5}y + \frac{11}{5} \Rightarrow b_{xy} = \frac{-1}{5}$$

Now $b_{yx} . b_{xy} = \frac{2}{15} \in [0, 1]$ which shows that the assumption is correct.

c. $r_{xy}^2 = b_{yx} . b_{xy} = \frac{2}{15} \Rightarrow r_{xy} = -\sqrt{\frac{2}{15}}$.

d. $\frac{b_{xy}}{b_{yx}} = \frac{r_{xy}\frac{s_x}{s_y}}{r_{xy}\frac{s_y}{s_x}} = \frac{s_x^2}{s_y^2} \Rightarrow \frac{s_x^2}{s_y^2} = \frac{3}{10}$.

Q2. Given

x:	15	25
Y:	50	30

Find the correlation coefficient between x and y.

Ans: Given only two points in scatter diagram, only one straight line can be drawn. So both the regression lines coincide on each other.

Hence $r_{xy} = \pm 1$

But from the given data the nature of change of x and y are opposite.

So $r_{xy} = -1$.

Q3. Marks of 5 students in Mathematics and Statistics are given below

Mathematics:	38	48	43	40	41
Statistics:	31	38	43	33	35

a. Determine the regression lines.
b. When the marks of a studet in Mathematics is 42, determine his most likely marks in Statistics.

Ans: Let x_i denotes the marks of i th studet in Mathematics

y_i denotes the marks of i th studet in Statistics

Assume the regression equation y on x be $y = a + bx$

The normal equations are $\sum y_i = na + b \sum x_i$ and $\sum x_i y_i = a \sum x_i + b \sum x_i^2$

Also assume the regression equation x on y be $x = c + dy$

The normal equations are $\sum x_i = nc + d \sum y_i$ and $\sum x_i y_i = c \sum y_i + d \sum y_i^2$

Take $u_i = x_i - 43$ and $v_i = y_i - 38$

So the normal equations become:

$\sum v_i = na + b \sum u_i$ and $\sum u_i v_i = a \sum u_i + b \sum u_i^2$ for v on u.

$\sum u_i = nc + b \sum v_i$ and $\sum u_i v_i = a \sum v_i + b \sum v_i^2$ for u on v.

x	y	u	v	uv	u^2	v^2
38	31	-5	-7	35	25	49
48	38	5	0	0	25	0
43	43	0	5	0	0	25
40	33	-3	-5	15	9	25
41	35	-2	-3	6	4	9
Total:		-5	-10	56	63	108

So from the normal equations: $-10 = 5a + (-5)b$

$$56= a(-5) +(63)b$$

Solving a=2.82 and b = 0.79

Hence the regression eqution y on x is y= 2.82 +0.79 x

And similarly solving the normal equations

C = 23.28 and d= 0.52

Hence the regression eqution x on y is

X= 23.28 + 0.52y

So when x=42, y = 2082 +0.79 (42) = 36.

Q4. The group of 10 workers in a factory is according to their efficiency two groups of different judges as follows:

Worker :	A	B	C	D	E	F	G	H	I	J
Judge1:	4	8	6	7	1	3	2	5	10	9
Judge2:	3	9	6	5	1	2	4	7	8	10

Ans; Taking u_i = rank of ith candidate given by judge 1

v_i = rank of ith candidate given by judge 2

Hence d_i: 1 -1 0 2 0 1 -2 -2 2 -1

d_i^2 : 1 1 0 4 0 1 4 4 4 1

$\sum_{i=1}^{10} d_i^2 = 19$

Hence $r_R = 1 - \frac{6\sum_{i=1}^{10} d_i^2}{n(n^2-1)} = 1 - \frac{6 X 19}{10X99} = 0.88$

Prepared by

Sanjay bhattacharya